









# Structural bases for aspartate recognition and polymerization efficiency of cyanobacterial cyanophycin synthetase

Received: 21 January 2022

Accepted: 8 August 2022

Published online: 30 August 2022

 Check for updates

Takuya Miyakawa <sup>1,7,8</sup>, Jian Yang <sup>1,2,3,8</sup>, Masato Kawasaki<sup>4,5,8</sup>,  
Naruhiko Adachi <sup>4,8</sup>, Ayumu Fujii<sup>1</sup>, Yumiko Miyauchi<sup>1</sup>, Tomonari Muramatsu<sup>1</sup>,  
Toshio Moriya <sup>4</sup>, Toshiya Senda <sup>4,5,6</sup>  & Masaru Tanokura <sup>1</sup> 

Cyanophycin is a natural biopolymer consisting of equimolar amounts of aspartate and arginine as the backbone and branched sidechain, respectively. It is produced by a single enzyme, cyanophycin synthetase (CphA1), and accumulates as a nitrogen reservoir during N<sub>2</sub> fixation by most cyanobacteria. A recent structural study showed that three constituent domains of CphA1 function as two distinct catalytic sites and an oligomerization interface in cyanophycin synthesis. However, it remains unclear how the ATP-dependent addition of aspartate to cyanophycin is initiated at the catalytic site of the glutathione synthetase-like domain. Here, we report the cryogenic electron microscopy structures of CphA1, including a complex with aspartate, cyanophycin primer peptide, and ATP analog. These structures reveal the aspartate binding mode and phosphate-binding loop movement to the active site required for the reaction. Furthermore, structural and mutational data show a potential role of protein dynamics in the catalytic efficiency of the arginine condensation reaction.

Bacteria efficiently convert carbon and nitrogen sources into various classes of intracellular and extracellular biopolymers with distinct chemical properties and biological functions<sup>1,2</sup>. These polymeric substances provide energy storage, adhesion, or protection functions in cells, and their production is regulated by environmental stimuli<sup>3</sup>. Over the past few decades, a better understanding of the molecular mechanism underlying the biosynthesis of biopolymers has shed light on the production of tailor-made biopolymers and their utility in the industrial field<sup>4,5</sup>. Among them, polypeptides contribute significantly

to a sustainable society due to their biomass origin, biodegradability, and unique functionality<sup>6</sup>.

Cyanophycin, a nonribosomally synthesized polypeptide consisting of equimolar amounts of aspartate and arginine as the backbone and branched sidechain, respectively, was first identified in cyanobacteria in the form of opaque and light-scattering cytoplasmic granules (Supplementary Fig. 1). Cyanophycin has high nitrogen to carbon ratio of 1:2, making it a reservoir of fixed nitrogen in most cyanobacteria<sup>7</sup>. Simultaneous oxygenic photosynthesis and N<sub>2</sub> fixation

<sup>1</sup>Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan. <sup>2</sup>CAS Key Laboratory of Tropical Marine Bio-resources and Ecology, Guangdong Key Laboratory of Marine Materia Medica, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China. <sup>3</sup>Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China. <sup>4</sup>Structural Biology Research Center, Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan. <sup>5</sup>Department of Materials Structure Science, School of High Energy Accelerator Science, The Graduate University of Advanced Studies (Soken-dai), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan. <sup>6</sup>Faculty of Pure and Applied Sciences, University of Tsukuba, 1-1-1 Tennodai, Ibaraki 305-8571, Japan. <sup>7</sup>Present address: Division of Integrated Life Science, Graduate School of Biostudies, Kyoto University, Kitashirakawa-oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan. <sup>8</sup>These authors contributed equally: Takuya Miyakawa, Jian Yang, Masato Kawasaki, Naruhiko Adachi.  e-mail: [toshiya.senda@kek.jp](mailto:toshiya.senda@kek.jp); [amtanok@mail.ecc.u-tokyo.ac.jp](mailto:amtanok@mail.ecc.u-tokyo.ac.jp)

are a significant challenge for microorganisms because the O<sub>2</sub> produced from CO<sub>2</sub> fixation is inhibitory to nitrogenase, which catalyzes the conversion of N<sub>2</sub> to NH<sub>3</sub><sup>8</sup>. To solve this problem, diazotrophs have developed physical strategies to temporally and spatially separate nitrogenase from O<sub>2</sub>. The heterocyst-forming cyanobacteria, such as *Anabaena* spp., perform CO<sub>2</sub> fixation and N<sub>2</sub> fixation in different cells: CO<sub>2</sub> fixation in the vegetative cells and N<sub>2</sub> fixation in heterocysts<sup>9</sup>. Other cyanobacteria, such as *Trichodesmium* spp., temporally segregate the processes by a diel cycle with CO<sub>2</sub> fixation during the day and N<sub>2</sub> fixation at night<sup>10</sup>. The fixed nitrogen accumulates at night as cyanophycin granules separated from the other cellular components but not as compounds that affect cellular metabolic dynamics. Cyanophycin is degraded by cyanophycinase and isopeptidase to release nitrogen to support cell growth during the day. Thus, cyanophycin granules serve as dynamic storage bodies in diazotrophic cyanobacteria to uncouple N<sub>2</sub> fixation from overall growth dynamics<sup>9</sup>. Cyanophycin accumulation enables the optimization of the nitrogen utilization of cyanobacteria in natural environments with fluctuating and limiting nitrogen supplies<sup>11</sup>.

The biosynthesis of cyanophycin is achieved by a single enzyme named cyanophycin synthetase (CphA1) that catalyzes peptide elongation coupled with ATP hydrolysis<sup>5</sup>. CphA1 uses a low-molecular-weight cyanophycin (at least 3–4 dipeptides long) as a primer for cyanophycin polymerization<sup>12</sup>. Recently, the tertiary structure of CphA1 has become available and reveals that amino acid polymerization is driven by three distinct domains<sup>13</sup>: the N-terminal domain (N domain), the middle glutathione synthetase-like domain (G domain) with the ATP-grasp fold, and the MurE-like muramyl ligase domain (M domain). The discrete active sites of the G and M domains are used for extending the Asp backbone and attaching the Arg sidechain, respectively (Supplementary Fig. 1). The N domain has been proposed to anchor cyanophycin polymers through electrostatic interactions<sup>13</sup>. However, it remains unclear how ATP-dependent addition of aspartate to cyanophycin is initiated at the catalytic site of the G domain.

In this work, given the importance of CphA1 in cyanophycin biosynthesis and the biotechnological value that structural information on enzyme-substrate interactions would provide, we investigate the CphA1 enzyme from the marine diazotroph *Trichodesmium erythraeum* IMS101 (*TeCphA1*), an important contributor to global nitrogen and carbon cycling<sup>14</sup>. We solve the *TeCphA1* structure bound to ATP analogs, cyanophycin primer peptides, and the substrate aspartate, in addition to apo and ATP analog-bound structures. Key elements, including residues and protein dynamics required for enzymatic activity, are identified based on experimental evidence of the G and M domains. In particular, we reveal the structural bases for aspartate recognition and condensation, which deepen our understanding of the reaction of cyanophycin biosynthesis in the G domain. Moreover, our data show a potential role of protein dynamics in the catalytic efficiency for the reaction of Arg sidechain condensation.

## Results

### Cryo-EM structures of *TeCphA1*

The *TeCphA1* structures in three different states were determined by single-particle cryogenic electron microscopy (cryo-EM): apo, the adenosine-5'-( $\gamma$ -thio)-triphosphate (ATPyS, an ATP analog)-bound state, and the substrate-bound state (*TeCphA1*-aspartate-ATPyS-cyanophycin primer peptide complex) at 2.91, 2.96, and 2.64 Å resolution, respectively (Fig. 1a, Supplementary Figs. 2–5, and Supplementary Table 1). The cryo-EM data of the substrate-bound state were collected under a high concentration of L-aspartate (100 mM), cyanophycin primer peptide, ( $\beta$ -Asp-Arg)<sub>4</sub>, and ATPyS. *TeCphA1* consists of three domains: an N-terminal domain (N domain, residues 1–161) that adopts a partially similar structure to *E. coli* RNA-polymerase  $\alpha$ -subunit<sup>13</sup>, a glutathione synthetase-like domain (G domain, residues 162–487), and a MurE-like muramyl ligase domain (M domain, residues

488–876) (Fig. 1b). *TeCphA1* basically adopts a tetrameric assembly with D<sub>2</sub> symmetry (Fig. 1a and Supplementary Fig. 6), which is a common architecture with another cyanobacterial CphA1 from *Synechocystis* sp. UTEX2470 (*SuCphA1*) (Supplementary Fig. 7)<sup>13</sup>.

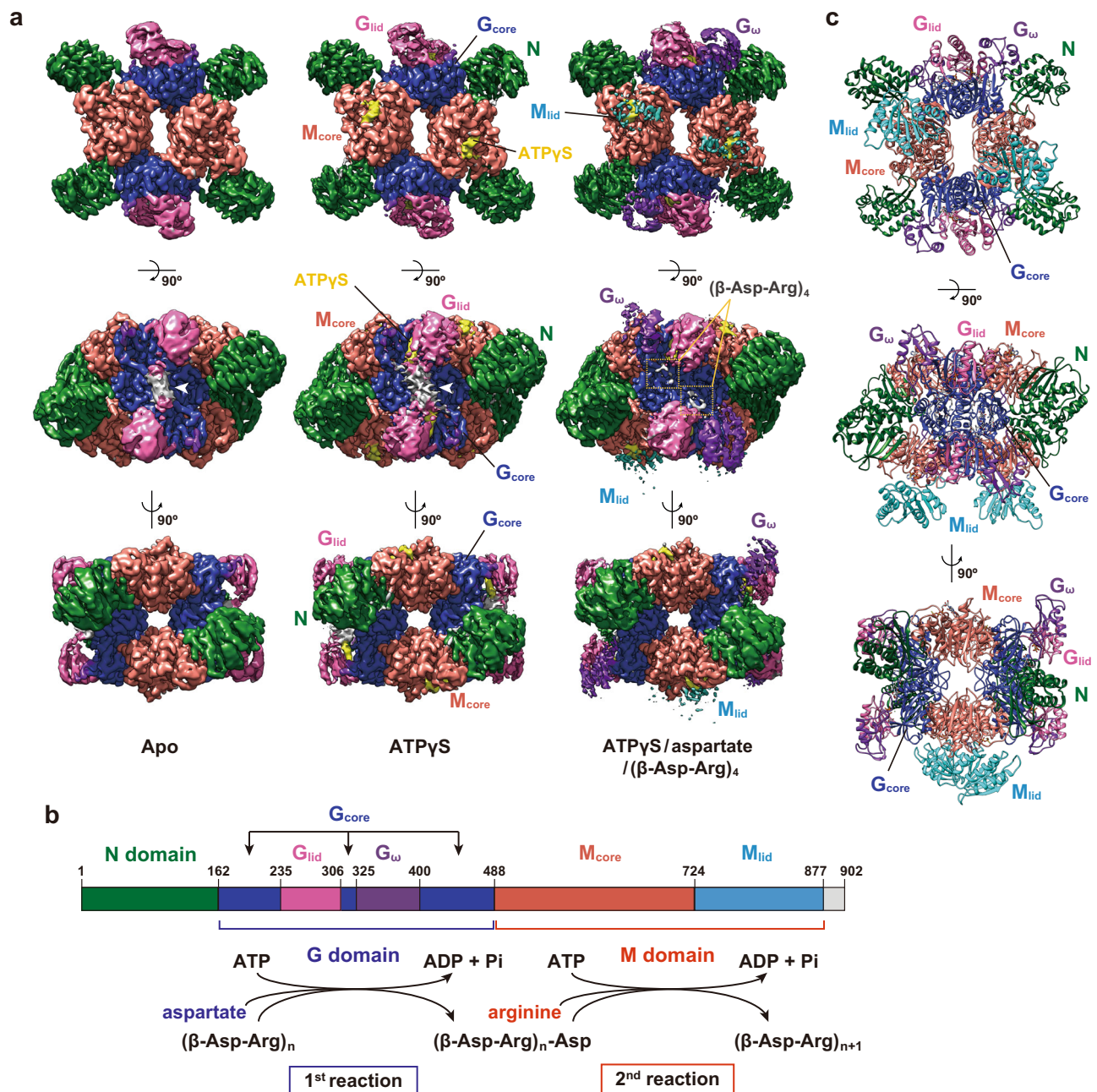
The G domain, which is responsible for backbone elongation through linking aspartate to the C-terminal backbone carboxy group of cyanophycin, is based on G<sub>core</sub> (residues 162–234, 306–324, and 400–487), with G<sub>lid</sub> (residues 235–305) and G<sub>ω</sub> (residues 325–399) modules inserted into the sequence (Fig. 1b and Supplementary Fig. 1). G<sub>core</sub> is a central module that provides a large dimer interface with another G<sub>core</sub> in the *TeCphA1* tetramer (Fig. 1c and Supplementary Fig. 8a). G<sub>lid</sub> acts as a lid of the ATP-binding site of G<sub>core</sub>, and G<sub>ω</sub> is positioned to surround the substrate-binding site together with G<sub>lid</sub>. The relative map intensity of G<sub>ω</sub> versus G<sub>core</sub> and G<sub>lid</sub> is higher in the substrate-bound state than in the other two states (Fig. 1a), suggesting that G<sub>ω</sub> is clearly visible as a result of reduced mobility, which is probably due to the binding of aspartate and ( $\beta$ -Asp-Arg)<sub>4</sub>.

The M domain, which catalyzes the addition of the Arg sidechain to the  $\beta$ -carboxy group of the C-terminal Asp residue of cyanophycin (Fig. 1b), is divided into two modules, M<sub>core</sub> (residues 488–723) and M<sub>lid</sub> (residues 724–876). M<sub>core</sub> contacts the intramolecular N domain and three G<sub>core</sub> to form the *TeCphA1* tetramer (Fig. 1c and Supplementary Fig. 8b), resulting in the clear density map of M<sub>core</sub> (Fig. 1a), and ATPyS binds to the surface pocket of M<sub>core</sub>. Therefore, M<sub>core</sub> assumes a distinct catalytic site from the G domain in *TeCphA1*, which is similar to the structural observation of *SuCphA1*<sup>13</sup>. On the other hand, the structural evidence of the ATP-bound forms of the G and M domains remains insufficient for the CphA1 enzymes from *Acinetobacter baylyi* DSM587 (*AbCphA1*) and *Tatumella morbirosei* DSM23827 (*TmCphA1*)<sup>13</sup>. M<sub>lid</sub> is not well resolved in the apo- and ATPyS-bound states (Fig. 1a and Supplementary Fig. 8c), suggesting the mobile nature of M<sub>lid</sub>. However, in the substrate-bound state, the M<sub>lid</sub> modules were partially resolved around ATPyS in chains A and B (Fig. 1a and Supplementary Fig. 6). The density of M<sub>lid</sub> was relatively weak compared with that of M<sub>core</sub>. A large shift of M<sub>lid</sub> was observed in the other CphA1 enzymes<sup>13</sup> and Mur ligases<sup>15–17</sup>.

*TeCphA1* has a high sequence identity (69.6%) and similarity (83.2%) to *SuCphA1* in the region consisting of the N, G, and M domains (residues 1–876). We compared the structures of *TeCphA1* in the substrate-bound state and *SuCphA1* that bound a cyanophycin analog ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> and an ATP analog 5'-adenylylmethylenediphosphonate (AMPPCP) to the G domain (PDB 7LGJ)<sup>13</sup>. These structures are similar to each other. The overall structures of the N domain and each module of the G and M domains (G<sub>core</sub>, G<sub>lid</sub>, G<sub>ω</sub>, M<sub>core</sub>, and M<sub>lid</sub>) were quite similar between *TeCphA1* and *SuCphA1* (root-mean-square deviation (RMSD) = 0.521–1.601 Å). The spatial arrangement of the N domain and the G<sub>core</sub> and M<sub>core</sub> modules was also well conserved between *TeCphA1* and *SuCphA1* (Supplementary Fig. 7b), whereas the relative orientation of the other modules versus G<sub>core</sub> and M<sub>core</sub> was different (Supplementary Fig. 7c, d). In particular, the distance between the M<sub>lid</sub> and N domain of *TeCphA1* is shorter than that of *SuCphA1* (Supplementary Fig. 7d).

### Functional roles of G<sub>core</sub> and G<sub>ω</sub> in adding aspartate to cyanophycin primer peptide

The *TeCphA1* structure in the substrate-bound state visualizes the initial state of the catalytic reaction for the ATP-dependent addition of aspartate to the C-terminal backbone carboxy group of cyanophycin (Fig. 1b and Supplementary Fig. 1). In the cryo-EM map processed with C<sub>1</sub> symmetry, all the *TeCphA1* protomers in the tetramer bound a cyanophycin primer peptide, ( $\beta$ -Asp-Arg)<sub>4</sub>, on the surface of G<sub>core</sub> (Fig. 1a). The cryo-EM map showed a clear density for the C-terminal  $\beta$ -Asp-Arg dipeptide unit (4<sup>th</sup> unit) and the other Asp backbones with a hook-like shape (Fig. 2a). The Arg sidechain of the 3<sup>rd</sup> unit partially showed weak density. Since the Arg sidechains of the 1<sup>st</sup> and 2<sup>nd</sup> units



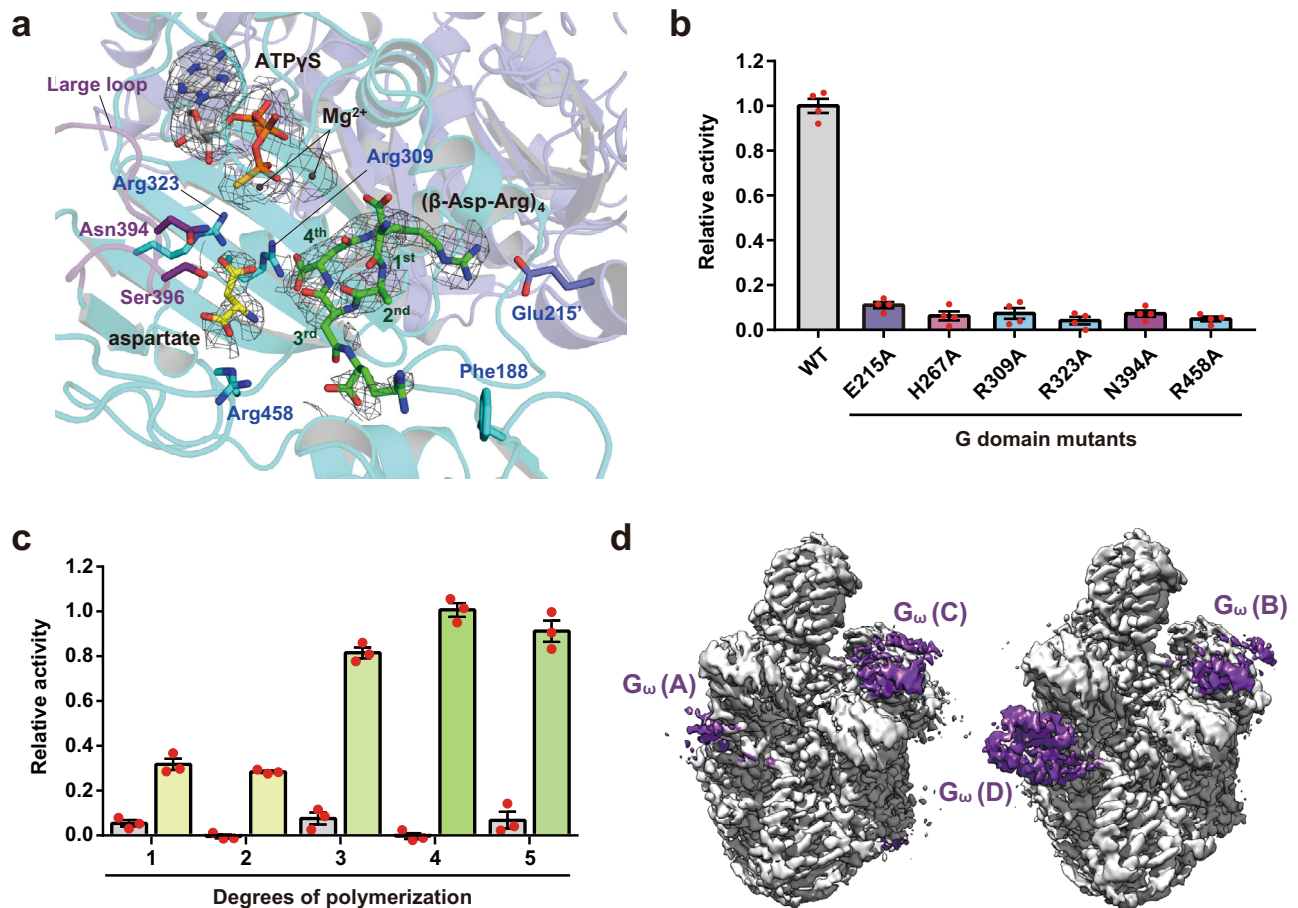
**Fig. 1 | Tetrameric structure and domain architecture of *TeCphA1*.** **a** Cryo-EM maps of *TeCphA1* in the apo, ATP $\gamma$ S-bound, and substrate (ATP $\gamma$ S/aspartate/( $\beta$ -Asp-Arg) $_4$ )-bound states. The maps are divided into six modules (N, G $_{core}$ , G $_{lid}$ , G $_{w}$ , M $_{core}$ , and M $_{lid}$ ) and ATP $\gamma$ S, which are shown in different color coordination. The map of ( $\beta$ -Asp-Arg) $_4$  is highlighted with a yellow dashed-line box in the density map of substrate-bound *TeCphA1*. A lump on the extra map (arrowhead) is observed

between two G $_{lid}$  modules in the apo and ATP $\gamma$ S-bound states. Each map is viewed from a twofold axis of tetrameric *TeCphA1*. **b** Residue range of each module on the *TeCphA* sequence and the reaction scheme of the G and M domains. The color coordination of each module is consistent with that in Panel **a**. **c** Ribbon diagram of the tetrameric *TeCphA* structure in the substrate-bound state. Each protomer is divided into the above-defined modules with the same color coordination as Panel **a**.

were invisible in the cryo-EM map, these models could not be built in the substrate-bound *TeCphA1* structure. In contrast, the Arg sidechain of the 2<sup>nd</sup> unit was visible in the cryo-EM map of *SuCphA1*<sup>13</sup>, and its guanidium group was located near Ala188. This residue is substituted with Phe on many cyanobacterial CphA1, including *TeCphA1* (Fig. 2a and Supplementary Fig. 9). The residue type at position 188 may affect the Arg-sidechain orientation of the 2<sup>nd</sup> unit on the active site of G $_{core}$ . In the *TeCphA1* structure, the C-terminal backbone carboxy group and Arg sidechain of the 4<sup>th</sup> unit interact with Arg309 (the nearest atomic distance, 2.9 Å) and Glu215, which is derived from another protomer forming the G $_{core}$ -G $_{core}$  dimer (Glu215') (2.8 Å), respectively (Fig. 2a and

Supplementary Fig. 8a). Since the catalytic activity of *TeCphA1* was impaired by the Ala substitutions of Glu215 and Arg309 (Fig. 2b and Supplementary Fig. 9), the interaction between the G $_{core}$ -G $_{core}$  dimer and the 4<sup>th</sup> unit seems to be required for the activity of *TeCphA1*.

The mutations of Glu215 and Arg309 (E215A and R309A) of the G domain were commonly analyzed in the two studies on *TeCphA1* and *SuCphA1*<sup>13</sup>. Of the two, R309A (and its corresponding mutation in *SuCphA1*) showed almost no activities. Both residues seem to be essential for the catalytic activities of *TeCphA1* and *SuCphA1*. On the other hand, while the assay conditions were not the same between the two, the E215A mutation was different between *TeCphA1* and *SuCphA1*.



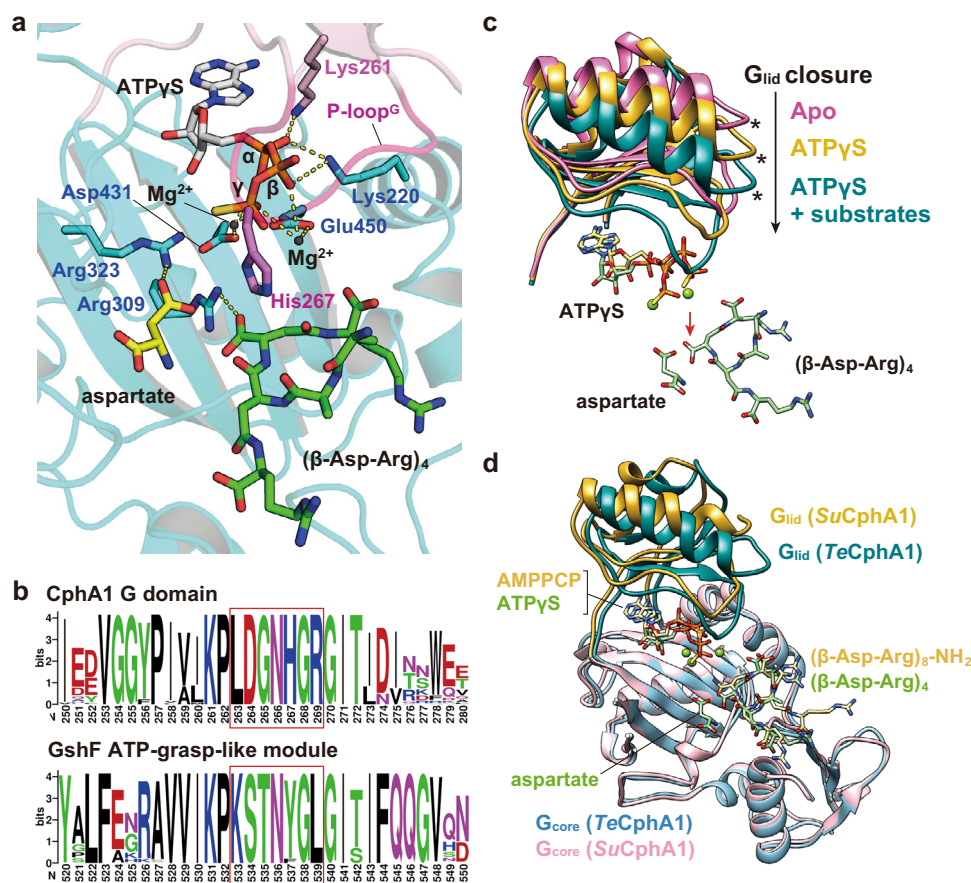
**Fig. 2 | Aspartate and cyanophycin primer peptide recognition at the G domain of *TeCphA1*.** **a** Binding site of ATP $\gamma$ S, the substrate aspartate, and a cyanophycin primer peptide, ( $\beta$ -Asp-Arg) $_4$ , on the surface of the G domain. The cryo-EM map carved 2 Å around Mg $^{2+}$ , ATP $\gamma$ S, the substrate aspartate, and ( $\beta$ -Asp-Arg) $_4$  at contour level 5. Individual chains of the G domain dimer are colored cyan (chain D) and medium blue (chain B). The purple structure shows the large loop of G $_{\omega}$  of chain D. The repeating units of ( $\beta$ -Asp-Arg) $_4$  are numbered from the N-terminus. **b** Relative enzymatic activity of *TeCphA1* (WT) and each G domain mutant, which was

measured by detecting released phosphates. Data are presented as mean  $\pm$  standard error of the mean (SEM) ( $n = 4$  independent experiments). **c** Relative activity of *TeCphA1* using the primer  $\beta$ -Asp-Arg peptide with different degrees of polymerization (DP = 1–5). The data of the reaction solution with (right) and without (left) *TeCphA1* are presented as mean  $\pm$  SEM ( $n = 3$  independent experiments). **d** Cryo-EM map of substrate-bound *TeCphA1*. Densities of the four G $_{\omega}$  modules of the *TeCphA1* tetramer are shown in purple.

The E215A of *TeCphA1* lost most of the catalytic activity (Fig. 2b), but the corresponding mutation of *SuCphA1* retained nearly half the activity. Since the interaction between E215 and the C-terminal Arg sidechain of cyanophycin contributes more to the cyanophycin binding in *TeCphA1* than in *SuCphA1*, the results of this mutational analysis seem to be explained by the difference in the structures. Although the importance of the Arg sidechains for the lack of poly-Asp polymerase activity has been explained by the structural observation of *SuCphA1*<sup>13,18</sup>, our results showed that the Arg sidechain of the 4<sup>th</sup> unit is a major contributor to ( $\beta$ -Asp-Arg) $_4$  recognition and determines the position of the 4<sup>th</sup> Asp backbone at the active site of the G domain to avoid polymerization progression with the aspartate backbone alone. Since ( $\beta$ -Asp-Arg) $_4$  is well accommodated at the binding site of G $_{\text{core}}$ , a cyanophycin molecule with at least 3–4 dipeptides long seems to be suitable as a primer peptide for *TeCphA1* activity (Fig. 2c). This observation is consistent with the previous results of the other CphA1<sup>18</sup>. According to the common structural features between *TeCphA1* and *SuCphA1*, the cyanophycin primer peptide seems to adopt a hook-like shape at the active site of CphA1 (Fig. 2a).

While ( $\beta$ -Asp-Arg) $_4$  was observed in all protomers, the substrate aspartate was observed only in one protomer (chain D) (Fig. 2a). The map intensity of G $_{\omega}$  was significantly different among the four protomers, and G $_{\omega}$  of chain D showed the highest map intensity of the four

G $_{\omega}$  (Fig. 2d). The substrate aspartate was visualized in the most stable G $_{\omega}$ . The other G $_{\omega}$  modules seemed to be more mobile than those in chain D. *TeCphA1* required four substrates, L-aspartate, L-arginine, ( $\beta$ -Asp-Arg) $_4$ , and ATP, for the catalytic reaction (Supplementary Fig. 10a). Steady-state kinetic analysis showed a typical Michaelis-Menten kinetics for L-arginine, ( $\beta$ -Asp-Arg) $_4$ , and ATP (Supplementary Fig. 10b). While we have determined the apparent  $K_m$  value for ATP, *TeCphA1* has two active sites in the G and M domains for the different catalytic reactions with ATP. As observed for the other CphA1 enzyme<sup>19</sup>, the sequential condensation reaction of *TeCphA1* may be rate-limited primarily by the site with lower affinity for ATP when the  $K_m$  values for ATP differ between the G and M domains. Unlike these substrates, aspartate acts as a positive effector of the catalytic reaction with the Hill coefficient ( $h$ )  $1.92 \pm 0.07$ . In addition,  $K_{\text{half}}$  and  $K_{\text{prime}}$  ( $= K_{\text{half}}^h$ ) were estimated as  $18.0 \pm 1.1$  mM and  $254 \pm 21$  mM, respectively. As shown in Supplementary Fig. 10b, the catalytic activity of *TeCphA1* increases with the increasing aspartate concentration; the cellular aspartate concentration may be a key regulator for the catalytic activity of *TeCphA1*. The results of the kinetic analysis for aspartate (Supplementary Fig. 10b) suggest that 100 mM aspartate is appropriate for the cryo-EM analysis. While it is unclear why the aspartate molecule was visible in only one subunit, the cryo-EM structure of *TeCphA1* with ATP $\gamma$ S, ( $\beta$ -Asp-Arg) $_4$ , and aspartate may be a



**Fig. 3** | G $_{lid}$  movement dependent on the binding of ATP and substrates, aspartate and ( $\beta$ -Asp-Arg) $_4$ . **a** Residue arrangement around ATP $\gamma$ S, the substrate aspartate, and the C-terminal backbone carboxy group of ( $\beta$ -Asp-Arg) $_4$ . Hydrogen bonds and Mg $^{2+}$  coordination are shown as yellow dashed lines. **b** Sequence logo depicting different conserved motifs between the P-loop $^G$  of CphA1 and the P-loop of the GshF ATP-grasp-like module. The P-loop sequences are shown in red boxes. The numbers under the logo correspond to the residue numbers of *TeCphA1* and

GshF from *Pasteurella multocida*. **c** Structural comparison of G $_{lid}$  among three states of *TeCphA1*. A comparison was performed after superposing the G $_{core}$  modules of the three states. Green spheres represent two Mg $^{2+}$  ions. Asterisks and red arrows indicate the position of Lys293 and the C-terminal backbone carboxy group of ( $\beta$ -Asp-Arg) $_4$ , respectively. **d** Superposed structures of the substrate-bound *TeCphA1* and the AMPPCP/( $\beta$ -Asp-Arg) $_5$ -NH $_2$ -bound *SuCphA1* structure (PDB 7LGJ) $^{13}$ .

snapshot of the equilibrium state of complex formation between the enzyme and aspartate.

The substrate aspartate was accommodated in a small space beside Ser396 (Fig. 2a). This binding mode seemed to be conserved in *SuCphA1* because *SuCphA1* activity was lost by the S396W mutation $^{13}$ . The substituted Trp residue prevented the reaction of Asp backbone elongation by occupying the space for substrate aspartate binding (Fig. 2a). Residues contributing to substrate aspartate recognition are conserved in the cyanobacterial CphA1 enzymes (Fig. 2a and Supplementary Fig. 9). Asn394, a residue on the large loop of G $_{\omega}$  (residues 389–399), and Arg323 interacted with the  $\beta$ -carboxy group of the substrate aspartate at distances of 2.5 Å and 3.4 Å, respectively (Fig. 2a). The  $\alpha$ -carboxy group of the substrate aspartate was close to the sidechain of Arg458 (2.5 Å) in the *TeCphA1* structure. Each Ala mutation of Asn394, Arg323 or Arg458 impaired *TeCphA1* activity (Fig. 2b). These observations are consistent with the notion that G $_{\omega}$  functions as a module for substrate aspartate recognition.

#### Catalytically important residues in G $_{core}$ and G $_{lid}$ for Asp backbone elongation

The G domain adopts an ATP-grasp fold that requires nucleophilic attack of the C-terminal backbone carboxy group of cyanophycin on ATP for Asp backbone elongation $^{13,20}$ . However, since these two functional groups are negatively charged, it is necessary to avoid

electrostatic repulsion between them. The ATP bound to the ATP-grasp fold typically interacts with one to three Mg $^{2+}$  ions $^{20}$ . In the G domain of *TeCphA1*, the  $\gamma$ -phosphate chelates two Mg $^{2+}$  ions together with two acidic residues, Asp431 and Glu450 (Figs. 2a and 3a). This structural feature is similar to *SuCphA1* $^{13}$  and other ATP-grasp enzymes, such as bacterial glutathione synthetases $^{21}$  and *N* $^5$ -carboxyaminoimidazole ribonucleotide synthetase (PurK) $^{20}$ . In addition, the  $\alpha$ - and  $\beta$ -phosphates of ATP $\gamma$ S interact with the positively charged sidechains of Lys261 and Lys220, respectively (Fig. 3a). While the triphosphate of ATP is neutralized overall by these interactions, there are still two conserved charged residues, Arg309 and Arg323, around the  $\gamma$ -phosphate. The substrate-bound *TeCphA1* structure shows that these residues electrostatically interact with the C-terminal backbone carboxy group of ( $\beta$ -Asp-Arg) $_4$  and  $\beta$ -carboxy group of the bound aspartate (Fig. 3a and Supplementary Fig. 9).

The substrate-bound-state structure suggests that the phosphate-binding loop of G $_{lid}$  (hereafter P-loop $^G$ , residues 263–269) (Fig. 3b) plays a critical role in the catalytic reaction. In the substrate-bound state, His267 on the P-loop $^G$  is located between  $\gamma$ -phosphate and the C-terminal backbone carboxy group of ( $\beta$ -Asp-Arg) $_4$  (Fig. 3a). This His267 position seems to be suitable to stabilize an acylphosphate intermediate together with Arg309, following the nucleophilic attack of the substrate aspartate. The importance of these residues was confirmed by the conservation of cyanobacterial CphA1 enzymes, including *SuCphA1* and decreased activity of a mutant substituted with

Ala both in *TeCphA1* and *SuCphA1* (Figs. 2b and 3b)<sup>13</sup>. This loop corresponds to the B loop of PurT and PurK<sup>22</sup>, which are ATP-grasp enzymes of *E. coli*. These enzymes function in the de novo purine biosynthetic pathway, and the B loop is required to efficiently generate an acylphosphate intermediate in these enzymes. However, His267, which seems to be important to stabilize the acylphosphate intermediate in *TeCphA1*, is not conserved in PurT and PurK<sup>22</sup>.

### **G<sub>lid</sub> motion in the catalytic mechanism of Asp backbone elongation**

Due to the charge neutralization of the  $\gamma$ -phosphate of ATP and the C-terminal backbone carboxy group of cyanophycin, these two groups can approach each other until reaching the distance of nucleophilic attack. A comparison of the three states of the *TeCphA1* structure (Fig. 3c) revealed conformational changes of G<sub>lid</sub> induced by ATP, the substrate aspartate and ( $\beta$ -Asp-Arg)<sub>4</sub> binding. These conformational changes seem to contribute to optimizing the relative dispositions of ATP, aspartate, and ( $\beta$ -Asp-Arg)<sub>4</sub> for the catalytic reaction. The binding of ATP $\gamma$ S causes G<sub>lid</sub> to move toward G<sub>core</sub> by 4.6 Å, which is the distance between the C $\alpha$  atoms of Lys293 (Fig. 3c, asterisk), and the substrate aspartate and ( $\beta$ -Asp-Arg)<sub>4</sub> binding induces a further shift of G<sub>lid</sub> by 4.3 Å. This G<sub>lid</sub> movement seems to change the binding position of ATP $\gamma$ S toward ( $\beta$ -Asp-Arg)<sub>4</sub>. Interestingly, G<sub>lid</sub> and ATP $\gamma$ S in the substrate-bound *TeCphA1* structure are located at a position closer to G<sub>core</sub> than those in the *SuCphA1* structure bound with the cyanophycin analog ( $\beta$ -Asp-Arg)<sub>8</sub>-NH<sub>2</sub> (Fig. 3d). Since the residues contributing to the reaction of Asp backbone elongation in the G domain are conserved between *TeCphA1* and *SuCphA1* (Supplementary Fig. 9), the binding of the substrate aspartate seems to modulate the orientation of ATP $\gamma$ S together with G<sub>lid</sub>.

While the binding of ATP, the substrate aspartate, and ( $\beta$ -Asp-Arg)<sub>4</sub> seems to cause the rearrangement of their relative orientations for the catalytic reaction, the  $\gamma$ -phosphate of ATP is less susceptible to nucleophilic attack by the C-terminal backbone carboxy group of ( $\beta$ -Asp-Arg)<sub>4</sub>, because they are located at a distance of  $\sim$ 6 Å (Fig. 3c). Although G<sub>lid</sub> moves toward the active site to change the relative positions of ATP and ( $\beta$ -Asp-Arg)<sub>4</sub>, a further mechanism is required to bridge this distance gap for the catalytic reaction to proceed. One of the possible mechanisms is the thermal motion of G<sub>lid</sub>, which is mentioned in relation to *SuCphA1*<sup>13</sup>. Since the cryo-EM structures of *TeCphA1* only represent the average position in the equilibrium of the G<sub>lid</sub> motion, it may be possible to consider that the protein dynamics allow these two molecules to occasionally access the distance of the catalytic reaction.

While the relative orientation of ATP $\gamma$ S and ( $\beta$ -Asp-Arg)<sub>4</sub> in the substrate-bound state needs a greater change for the catalytic reaction, the relative orientation of ( $\beta$ -Asp-Arg)<sub>4</sub> and the substrate aspartate in the G domain seems to be suitable for ATP-dependent backbone elongation (Fig. 1b and Supplementary Fig. 1). The amino group of substrate aspartate is located 4.1 Å from the carboxy group at the C-terminus of ( $\beta$ -Asp-Arg)<sub>4</sub>, and the position of the carboxy group may be adjusted by phosphorylation (Figs. 2a and 3c).

### **Structural and functional roles of M<sub>lid</sub> in Arg sidechain condensation**

In the substrate-bound state, M<sub>core</sub> was clearly visible in the four protomers of the *TeCphA1* tetramer. ATP $\gamma$ S interacts with the conserved P-loop of M<sub>core</sub> (hereafter P-loop<sup>M</sup>, residues 495–501) (Fig. 4a and Supplementary Fig. 9). Lys499 and an Mg<sup>2+</sup> ion are located between the  $\beta$ -phosphate and  $\gamma$ -thiophosphate groups of ATP $\gamma$ S, and the Mg<sup>2+</sup> ion is chelated by Thr500, Thr522, and Glu558. These interactions with ATP are highly conserved in the Mur ligases, which synthesize the peptide stem of bacterial peptidoglycan<sup>23,24</sup>. Since the Ala mutation of Lys499 impaired *TeCphA1* activity (Fig. 4b), the interaction between Lys499 and ATP is also critical to the catalytic activity of CphA1 enzymes.

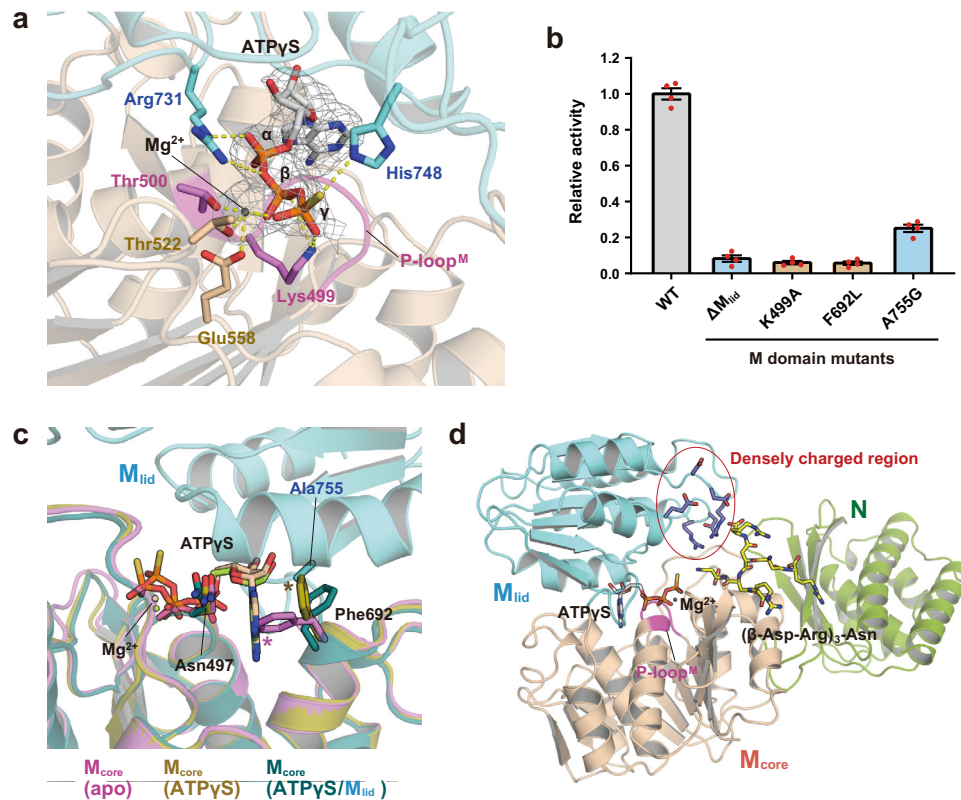
While ATP $\gamma$ S is visible in the M<sub>core</sub> modules of all protomers, the cryo-EM densities for M<sub>lid</sub> were observed weakly only in protomers A and B (Supplementary Fig. 11), suggesting a relatively mobile nature of M<sub>lid</sub>, even in the substrate-bound state. In protomers A and B, M<sub>lid</sub> and M<sub>core</sub> sandwich ATP $\gamma$ S (Figs. 1a and 4a), and Arg731 and His748 in M<sub>lid</sub> interact with the  $\alpha$ -phosphate and  $\gamma$ -thiophosphate groups of ATP $\gamma$ S. Since M<sub>lid</sub> was not observed in the ATP $\gamma$ S-bound state, M<sub>lid</sub> is unlikely essential to the stable binding of ATP $\gamma$ S.

However, since the deletion mutant of M<sub>lid</sub> ( $\Delta$ M<sub>lid</sub>) showed substantially reduced activity of *TeCphA1*, the M<sub>lid</sub> module is necessary for the efficient catalytic reaction of *TeCphA1* (Fig. 4b) and seems to play a significant role in ATP-dependent Arg sidechain condensation. Superposition of *TeCphA1* in apo- and ATP $\gamma$ S-bound structures reveals a conformational change of Phe692 in M<sub>core</sub> upon ATP $\gamma$ S binding, and Phe692 and Asn497 sandwich the adenine ring of ATP $\gamma$ S (Fig. 4c). This sidechain orientation of Phe692 causes steric hindrance with the methyl group of Ala755 in M<sub>lid</sub> (Fig. 4c and Supplementary Fig. 12). On the other hand, the sidechain position of Phe692 in the apo form inhibits ATP binding but not the interaction between M<sub>core</sub> and M<sub>lid</sub> (Fig. 4c). The flexible nature of M<sub>lid</sub> and the conformational change of Phe692 seem to cause mutually exclusive binding conformations of ATP and M<sub>lid</sub>, which may facilitate the exchange of ATP and ADP in the reaction cycle. The invisible M<sub>lid</sub> in protomers C and D can be explained by mutually exclusive binding conformations, and protomers C and D are considered ATP $\gamma$ S-binding states. Intriguingly, protomers A and B of the substrate-bound state showed simultaneous binding of ATP $\gamma$ S and M<sub>lid</sub> to M<sub>core</sub>. We consider that M<sub>lid</sub> and ATP $\gamma$ S in protomers A and B are in a metastable state that can be realized due to a high concentration of ATP $\gamma$ S. The interaction between Ala755 and Phe692 in protomers A and B may be energetically unfavored and thus easily changed to the M<sub>lid</sub>-binding or ATP $\gamma$ S-binding state under lower ATP $\gamma$ S concentrations.

While ATP $\gamma$ S was visible on all M domains, ( $\beta$ -Asp-Arg)<sub>4</sub> was not observed on the M domain of the substrate-bound *TeCphA1* structure. This observation is consistent with the previous result that the Arg-unmodified Asp residue at the C-terminus of cyanophycin is required to bind to the M domain in the *SuCphA1* structure<sup>13</sup>. When the *TeCphA1* structure is superposed with the *SuCphA1* structure in complex with the cyanophycin analog ( $\beta$ -Asp-Arg)<sub>8</sub>-Asn<sup>13</sup>, the cyanophycin analog is located between the N domain and M<sub>core</sub> of *TeCphA1*. Interestingly, we found a densely charged region with multiple Arg and Asp/Glu residues on the surface of M<sub>lid</sub> (Fig. 4d). These residues seem to interact with the cyanophycin analog, as observed in *SuCphA1*. M<sub>lid</sub> may contribute to guiding cyanophycin into the active site of M<sub>core</sub>. In the Mur ligases, the interaction with the substrate induces a rotation of the C-terminal domain, which corresponds to M<sub>lid</sub> of CphA1 enzymes, resulting in a closed conformation<sup>15,17</sup>. However, the orientation of M<sub>lid</sub> is almost unchanged by the substrate (cyanophycin) interaction in *SuCphA1* (RMSD = 0.158 Å, 2,583 atoms of M<sub>core</sub> and M<sub>lid</sub>)<sup>13</sup>.

## **Discussion**

Cyanophycin is a unique biopolymer that consists of the  $\beta$ -Asp-Arg dipeptide as a repeating unit (Supplementary Fig. 1) and is produced by CphA1-dependent peptide synthesis. Our understanding of this polymerization process has been greatly improved by a recent report that provided structural snapshots of CphA1<sup>13</sup>. However, the catalytic and regulatory mechanisms of this reaction have not been fully elucidated. In the present study, we determined the CphA1 structure with a complete set of substrates for the reaction of the G domain, aspartate and ( $\beta$ -Asp-Arg)<sub>4</sub>, which visualizes the initial state for ATP-dependent addition of aspartate to the C-terminal backbone carboxy group of a cyanophycin primer peptide. This structure proposes a feasible catalytic mechanism for the Asp backbone elongation reaction that occurs in the G domain (Fig. 5). In the apo state, G<sub>lid</sub> adopts an open conformation and is the furthest away from G<sub>core</sub>. ATP binding induces the movement of G<sub>lid</sub> toward G<sub>core</sub> to sandwich an ATP molecule, whereas



**Fig. 4 | Structural bases of  $M_{lid}$  for ATP-dependent Arg sidechain condensation to cyanophycin.** **a** Binding site of ATP $\gamma$ S between  $M_{core}$  (wheat) and  $M_{lid}$  (cyan) on the surface of the M domain. The cryo-EM map carved 2 Å around  $Mg^{2+}$  and ATP $\gamma$ S at contour level 10. The magenta structure shows the P-loop $^M$ . Hydrogen bonds and  $Mg^{2+}$  coordination are shown as yellow dashed lines. **b** Relative enzymatic activity of *TeCphA1* (WT) and each M domain mutant, including the deletion mutant of  $M_{lid}$  ( $\Delta M_{lid}$ ). Enzyme activities were measured by detecting the released phosphates. Data are presented as mean  $\pm$  standard error of the mean (SEM) ( $n = 4$  independent

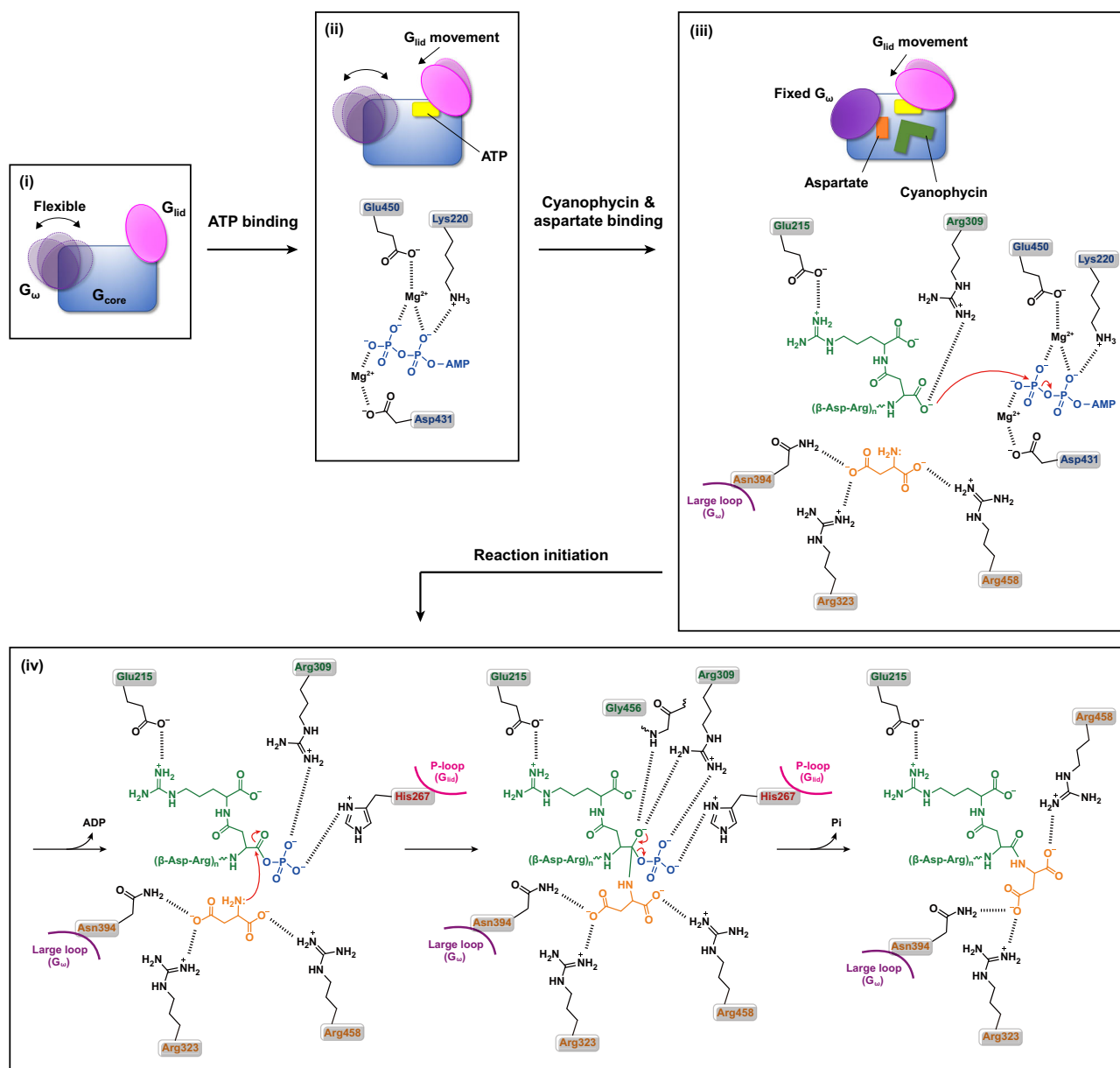
experiments). **c** Superposed structures of the M domain in three *TeCphA1* states. Asterisks (\*) represent the steric clashes of Phe692 with the adenine ring of ATP $\gamma$ S (pink) and Ala755 (ocher). **d** Localization of a cyanophycin analog, ( $\beta$ -Asp-Arg) $_3$ -Asn, on the N and M domains of *TeCphA1*. The model of ( $\beta$ -Asp-Arg) $_3$ -Asn was created by the superposition between the  $M_{core}$  modules of the substrate-bound *TeCphA1* structure and the ATP/( $\beta$ -Asp-Arg) $_8$ -Asn-bound *SuCphA1* structure (PDB 7LGQ) $^{13}$ .

$G_{\omega}$  remains flexible. Cyanophycin and aspartate binding further changes  $G_{lid}$  to a closed conformation, and the orientation of  $G_{\omega}$  is fixed to localize aspartate in a suitable position for the reaction by the interaction of Asn394 on the large loop with the substrate. Thus, ATP, the substrate aspartate, cyanophycin, and residues, including His267 on the P-loop $^C$ , are arranged to initiate the reaction of Asp backbone elongation. The C-terminal backbone carboxy group of cyanophycin is activated by ATP to form a high-energy acylphosphate intermediate required for the ATP-grasp enzymes $^{22}$ , which is stabilized by His267 on the P-loop $^C$  together with Arg309. The amino group of the substrate aspartate nucleophilically targets the carboxy carbon of the acylphosphate, thus generating a tetrahedral intermediate in which an oxyanion is stabilized by the interaction with Arg309 and Gly456. A new peptide bond forms between cyanophycin and the substrate aspartate with the release of phosphate.

The G domain of CphA1 enzymes adopts a fold similar to the ATP-grasp-like module of bacterial hybrid-type glutathione synthetase, GshF, which catalyzes peptide bond formation between the amino group of glycine and the C-terminal backbone carboxy group of  $\gamma$ -glutamyl cysteine $^{13,25}$ . The amino acid sequence of the P-loop $^C$  in CphA1 is different from that in GshF. In particular, the P-loop $^C$  has a His residue (His267) in CphA1 but a Tyr or Phe residue in GshF (Fig. 3b). In addition, Glu215 and Arg458 of *TeCphA1* are substituted with Leu and Met/Tyr in GshF, respectively. These different residues of CphA1 from GshF seem to be optimized for the catalytic reaction for the Asp backbone elongation of cyanophycin. Mutational analysis suggested that His267 plays a critical role in the catalytic reaction for *TeCphA1*

(Figs. 2b and 5). Glu215 and Arg458 are required for the recognition of ( $\beta$ -Asp-Arg) $_4$  and the substrate aspartate at the active site of  $G_{core}$ , respectively (Fig. 2a, b).

Although Tlr2120 from *Thermosynechococcus elongatus* BP-1 has been found to be the only CphA1 enzyme that synthesizes cyanophycin in a primer-independent manner $^{26}$ , the other known CphA1 enzymes show primer-dependent activity. Efficient CphA1 activity requires a cyanophycin primer at least 3–4 dipeptides long (Fig. 2c) $^{18}$ . Hence, to rapidly initiate cyanophycin synthesis upon the transition to the  $N_2$  fixation phase, cyanophycin primer peptides must be available in cyanobacterial cells. Another critical factor for CphA1 activity is the cellular level of aspartate. In the process of  $N_2$  fixation, aspartate is produced by the incorporation of nitrogen into oxaloacetate and is consumed directly for cyanophycin synthesis $^9$ . Aspartate is also required as a nitrogen donor for arginine biosynthesis in cyanobacteria $^{27}$ . Therefore, the intracellular aspartate concentration is one of the factors most likely to affect the progression of cyanophycin synthesis. The steady-state kinetics of *TeCphA1* suggests an allosteric effect with positive cooperativity toward aspartate, and the reaction is, in fact, accelerated with an increasing aspartate concentration up to quite a high concentration range (Supplementary Fig. 10b). The structural comparison between the substrate-bound *TeCphA1* and the *SuCphA1* bound with the cyanophycin analog ( $\beta$ -Asp-Arg) $_8$ -NH $_2$  showed that the binding of aspartate modulates the orientation of ATP together with  $G_{lid}$  (Fig. 3d). This aspartate-mediated regulation of  $G_{lid}$  seems reasonable to control the CphA1 activity associated with storage of fixed nitrogen in cyanobacteria to avoid wasteful phosphorylation



**Fig. 5 | Schematic diagram of the catalytic reaction for aspartate addition to the C-terminal backbone carboxy group of cyanophycin.** (i) G domain from a protomer in the apo state. Three modules,  $G_{\text{core}}$ ,  $G_{\text{lid}}$ , and  $G_{\omega}$ , are shown in blue, pink, and purple, respectively. (ii) ATP-bound state. The chemical structure of ATP is depicted as AMP diphosphate. (iii) Substrate-bound state. The orientation of  $G_{\text{lid}}$  further changes compared to the ATP-bound state. (iv) Feasible catalytic

mechanism in the G domain. In this mechanism, a tetrahedral intermediate is formed by nucleophilic attack to an acylphosphate intermediate of cyanophycin by the amino group of aspartate. Arg309 is assumed to stabilize the acylphosphate intermediate and the tetrahedral intermediate together with His267 and Gly456, respectively.

of primer peptides under conditions in which aspartate is not available to CphA1 as a substrate.

In the apo and ATP $\gamma$ S-bound states, a large density was observed between two  $G_{\text{lid}}$  modules and over the binding site of  $(\beta\text{-Asp-Arg})_4$  (Fig. 1a, arrowhead). CphA1 is known to synthesize cyanophycin in *E. coli* cells<sup>28</sup>; hence, cyanophycin is the most likely candidate for the molecule with increased density. In addition, the polymerization reaction of CphA1 often terminates up to a polymer length of 25–30 kDa<sup>26</sup>. The mobility of  $G_{\text{lid}}$  is likely to be affected by the association of a high-molecular-weight molecule into the space between two  $G_{\text{lid}}$  modules, and the regulation of  $G_{\text{lid}}$  motion may also be related to the length determination and polymerization termination of cyanophycin, although this mechanism is a question that needs to be addressed in

further study. On the other hand, the windshield wiper model using the N domain has been proposed as a promising model for predicting the efficiency of reactions using two distant active sites alternately<sup>13</sup>. CphA2, which synthesizes cyanophycin by directly linking  $\beta\text{-Asp-Arg}$  dipeptides, lacks not only the P-loop<sup>M</sup> but also the  $M_{\text{lid}}$  module to abolish the activity for Arg sidechain condensation<sup>29</sup>, which is supported by our mutational data of *TeCphA1*, which does not exert the CphA2 activity (Fig. 4b and Supplementary Fig. 10a). This molecular evolution may imply that the presence of  $M_{\text{lid}}$  is unfavorable for the reaction to proceed efficiently only in the G domain, rather than alternately in the two domains. The mobility of  $M_{\text{lid}}$  and its densely charged region (Fig. 4c) may play a role in the reactivity of both active sites together with the N domain.



## Methods

### Materials

Cyanophycin primer peptides  $\beta$ -Asp-Arg,  $(\beta$ -Asp-Arg)<sub>2</sub>,  $(\beta$ -Asp-Arg)<sub>3</sub>,  $(\beta$ -Asp-Arg)<sub>4</sub>, and  $(\beta$ -Asp-Arg)<sub>5</sub> were synthesized via the solid phase method by Sangon Biotech (Shanghai, China). The synthesized peptides were confirmed using a TripleTOF 5600+ system (Sciex) with a TurbolonSpray (TIS) probe (Turbo V ion source, Sciex). Time of flight-mass spectroscopy (TOF-MS) spectra ( $m/z$  range, 100–2000 Da) were recorded in positive-ion mode with the following parameters: curtain gas, 15 psi; ion source gas 1, 15 psi; ion source gas 2, 0 psi; ion spray floating voltage, 5500 V; and interface heater temperature, 400 °C. The mass spectrometer was recalibrated using an APCI positive calibration solution (Sciex). The obtained spectra are listed in Supplementary Fig. 13. All the other chemicals were purchased from Sigma-Aldrich or FUJIFILM Wako Pure Chemical Corporation.

### Expression and purification of *TeCphA1*

The gene encoding *TeCphA1* (ENA code: ABG51217) was amplified from the genomic DNA of *T. erythraeum* IMS101 with forward and reverse primers (CphA-for: TATACATATGAAATCCTCAAACCTCCAGA CAT-TACG; CphA-rev: GGTGCTCGAGAATTGAACTTTTAAAACCTT) containing 5' 10-bp overhangs complementary to the vector sequence for seamless cloning (TransGen Biotech). The target gene was cloned into the pET-22b(+) vector (Novagen) between the restriction sites *NdeI* and *XhoI* along with a C-terminal hexa-histidine tag. All *TeCphA1* mutants were prepared by inverse PCR for site-specific mutagenesis using PrimeSTAR Max DNA polymerase (Takara Bio) and primers listed in Supplementary Table 2.

Recombinant proteins were expressed in the *E. coli* strain BL21(DE3) (Novagen). Cells were inoculated in LB media supplemented with 50 mg ml<sup>-1</sup> ampicillin and cultured overnight. Upon reaching an optical density of ~0.6 at 600 nm, the temperature was decreased from 37 °C to 16 °C, and the culture was supplemented with 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside. Protein expression was performed for 16 h, after which cells were harvested by centrifugation at 5,180  $\times g$  for 10 min, followed by resuspension in lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM dithiothreitol (DTT), and 5 mM imidazole). Cells were lysed by sonication, and cell debris was then removed by centrifugation at 40,000  $\times g$  for 30 min. The supernatant fraction was loaded onto an equilibrated Ni-nitrilotriacetic acid (NTA) resin (Qiagen). After extensive washing with wash buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, and 20 mM imidazole), the recombinant proteins were eluted with elution buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1 mM DTT, and 200 mM imidazole). The eluate was further purified by loading onto a Mono Q column (Cytiva) for anion exchange chromatography, and elution was then performed with a linear gradient of 0 to 1.0 M NaCl. After concentration with Vivaspin 20 devices (30 kDa cutoff; Sartorius), the protein solution was loaded onto a Superdex200 10/300 GL column (Cytiva) for size exclusion chromatography with buffer containing 20 mM Tris-HCl (pH 8.0), 400 mM NaCl, and 1 mM DTT. The purified *TeCphA1* and its mutants were concentrated with Vivaspin 20 devices and then stored in buffer for size exclusion chromatography at -80 °C until use.

### Grid preparation and cryo-EM data collection

Cryo-EM samples were prepared by mixing 24  $\mu$ M *TeCphA1* with each solution. The final sample compositions were as follows: (1) apo state, 12  $\mu$ M *TeCphA1* in a solution containing 20 mM Tris-HCl (pH 8.0), 400 mM NaCl, and 1 mM DTT; (2) ATP $\gamma$ S-bound state, 2  $\mu$ M *TeCphA1* in a solution containing 50 mM Tris-HCl (pH 8.2), 200 mM NaCl, 20 mM KCl, 20 mM MgCl<sub>2</sub>, 1 mM DTT, 4 mM ATP $\gamma$ S, 5 mM L-aspartate, 0.5 mM L-arginine, and 10 mM  $(\beta$ -Asp-Arg)<sub>4</sub>; and (3) substrate-bound

state, 4  $\mu$ M *TeCphA1* in a solution containing 50 mM Tris-HCl (pH 8.2), 67 mM NaCl, 20 mM KCl, 20 mM MgCl<sub>2</sub>, 1 mM DTT, 10 mM ATP $\gamma$ S, 100 mM L-aspartate, and 20 mM  $(\beta$ -Asp-Arg)<sub>4</sub>. Three microliters of sample was applied to a holey carbon grid (Quantifoil, Cu, R1.2/1.3, 300 mesh) rendered hydrophilic by a 30 s glow-discharge in air (11 mA current with PIB-10). The grid was blotted for 15 s with a blot force of 0 (sample 1), for 15 s with a blot force of 10 (sample 2), and for 10 s with a blot force of 10 (sample 3) and flash-frozen in liquid ethane using Vitrobot Mark IV (Thermo Fisher Scientific) at 18 °C and 100% humidity.

For apo and ATP $\gamma$ S-bound states, cryo-EM datasets were acquired with a Talos Arctica (Thermo Fisher Scientific) transmission electron microscope (TEM) operating at 200 kV in nanoprobe mode using EPU software for automated data collection (Supplementary Fig. 5a). The movie frames were collected by a 4 k  $\times$  4 k Falcon 3 direct electron detector (DED) in electron counting mode at a nominal magnification of 120,000, which yielded a pixel size of 0.88 Å pixel<sup>-1</sup>. Forty-nine and fifty movie frames were recorded at an exposure of 1.02 and 1.00 electrons per Å<sup>2</sup> per frame, corresponding to a total exposure of 50 e Å<sup>-2</sup>, respectively. The defocus range was -1.0 to -2.5  $\mu$ m for the apo and ATP $\gamma$ S-bound states. The dataset for the substrate-bound state was obtained with a 300 kV Titan Krios G3i TEM (Thermo Fisher Scientific) equipped with a Gatan Imaging Filter (GIF), Quantum energy filter (Gatan), and a K3 Summit DED (Gatan) (Supplementary Fig. 5a). The dataset was acquired in counting mode with a nominal magnification of 105,000 (yielding a calibrated pixel size of 0.83 Å pixel<sup>-1</sup>) using SerialEM software<sup>30</sup>. Forty-nine movie frames were recorded at an exposure of 1.00 electrons per Å<sup>2</sup> per frame, corresponding to a total exposure of 49 e Å<sup>-2</sup> and -0.84 to -1.8  $\mu$ m of defocus range. Data collection details are listed in Supplementary Table 1.

### Cryo-EM data processing

For the apo state dataset, 2,109 acquired movies were subjected to beam-induced motion correction by MotionCorr2<sup>31</sup>. The contrast transfer function (CTF) parameters were calculated using Gctf<sup>32</sup>. A total of 340,000 particles were fully automatically selected using SPHIRE crYOLO<sup>33,34</sup> with a generalized model using a selection threshold of 0.1. The subsequent processes were performed using RELION3.1<sup>35</sup>. The particles were extracted by downsampling to a pixel size of 2.64 Å pix<sup>-1</sup> and subjected to 2D classification. A total of 180,018 particles (17 classes) that displayed secondary-structural elements were selected for ab initio map reconstruction.  $D_2$  symmetry was imposed on the generated ab initio map and used as an initial reference map for the 3D classification assuming asymmetry ( $C_1$ ). A total of 103,041 particles associated with the best 3D class with the highest resolution were recentered, re-extracted with a pixel size of 0.88 Å pix<sup>-1</sup> and subjected to 3D refinement with  $D_2$  symmetry. Three rounds of CTF refinement<sup>35</sup> and Bayesian polishing<sup>36</sup> followed by 3D refinement with a soft-edged 3D mask yielded a  $D_2$  map with 3.11 Å resolution. To further improve the resolution, additional particles were selected using SPHIRE crYOLO<sup>33,34</sup> with a relaxed selection threshold of 0.02. A total of 556,730 particles were extracted by downsampling to a pixel size of 2.64 Å pix<sup>-1</sup> and subjected to 2D classification. A total of 429,149 particles (14 classes) that displayed secondary-structural elements were selected. The  $D_2$  map with a resolution of 3.11 Å obtained above was used as a reference map for the 3D classification assuming asymmetry ( $C_1$ ). A total of 426,572 particles associated with the best three 3D classes were recentered and re-extracted with a pixel size of 0.88 Å pix<sup>-1</sup> and subjected to 3D refinement with  $D_2$  symmetry. Three rounds of CTF refinement<sup>35</sup> and Bayesian polishing<sup>36</sup> followed by 3D refinement with a soft-edged 3D mask yielded a 3.03 Å resolution map. Then, no-alignment 3D classification was conducted ( $D_2$  symmetry, two expected classes,  $T=8$ ) with a soft-edged 3D mask, and 161,823 particles were selected by

choosing the best 3D class. The last 3D refinement ( $D_2$  symmetry, with a mask diameter of 260 Å) with a soft-edged 3D mask and subsequent postprocessing yielded a map with a global resolution of 2.91 Å according to the Fourier shell correlation (FSC) = 0.143 criteria<sup>37</sup>. The details of the cryo-EM data processing are shown in Supplementary Fig. 2, and the associated parameters and statistics are summarized in Supplementary Table 1. Directional FSC plots were calculated using the 3DFSC server to evaluate the evenness of the distribution of particle orientations (Supplementary Fig. 5b, c)<sup>38</sup>.

For the dataset of the ATP $\gamma$ S-bound state, 1,730 acquired movies were subjected to beam-induced motion correction by MotionCor2<sup>30</sup>. The CTF parameters were calculated using Gctf<sup>32</sup>. A total of 927,603 particles were picked fully automatically using SPHIRE crYOLO<sup>33,34</sup> with a generalized model using a selection threshold of 0.001. The subsequent processes were performed by using RELION3.1<sup>35</sup>. The particles were extracted by downsampling to a pixel size of 2.64 Å pix<sup>-1</sup> and subjected to 2D classification. The 424,332 particles (16 classes) that displayed secondary-structural elements were selected for ab initio map reconstruction.  $D_2$  symmetry was imposed on the generated ab initio map and used as an initial reference map for the 3D classification assuming asymmetry ( $C_1$ ). A total of 210,971 particles associated with the best 3D class with the highest resolution were recentered and re-extracted with a pixel size of 0.88 Å pix<sup>-1</sup> and subjected to 3D refinement with  $D_2$  symmetry. Two rounds of CTF refinement<sup>35</sup> and Bayesian polishing<sup>36</sup> followed by 3D refinement with a soft-edged 3D mask yielded a  $D_2$  map with 2.91 Å resolution. Then, no-alignment 3D classification was conducted ( $D_2$  symmetry, two expected classes,  $T = 16$ ) with a soft-edged 3D mask, and 49,842 particles were selected by choosing the best 3D class. The last 3D refinement ( $D_2$  symmetry, with a mask diameter of 260 Å) with a soft-edged 3D mask and subsequent postprocessing yielded a map with a global resolution of 2.96 Å according to the FSC = 0.143 criteria<sup>37</sup>. The processing strategy is shown in Supplementary Fig. 3, and the associated parameters and statistics are summarized in Supplementary Table 1. Directional FSC plots were calculated using the 3DFSC server to evaluate the evenness of the distribution of particle orientations (Supplementary Fig. 5b, c)<sup>38</sup>.

For the datasets of substrate-bound states, all 3,843 acquired movies were dose-fractionated and subjected to beam-induced motion correction implemented in RELION 3.1<sup>35</sup>, and 3,735 motion-corrected micrographs were selected with total motion < 34 Å. The CTF parameters were calculated using CTFFIND4<sup>39</sup>, and 3,730 micrographs were then selected with the estimated maximum resolutions < 6 Å. A total of 1,371,057 particles were automatically selected from the selected micrographs in RELION 3.1<sup>35</sup> using template images, which were generated from 1,338 particles (9 classes) by 2D classification of the 2,008 manually selected particles. The particles were extracted by downsampling to a pixel size of 2.49 Å pix<sup>-1</sup> and subjected to several rounds of 2D classification. The 296,927 particles (77 classes) were selected from the last round of 2D classification and applied to two rounds of 3D classification assuming asymmetry ( $C_1$ ). A total of 195,538 particles associated with the best 3D class with the highest resolution of the second round were recentered and re-extracted with a pixel size of 0.83 Å pix<sup>-1</sup> and subjected to 3D refinement. CTF refinement<sup>35</sup> and Bayesian polishing<sup>36</sup> were applied to the resulting particles to refine the per-particle defocus, beam-tilt, and beam-induced motion corrections. The map generated by the 3D refinement with a soft-edged 3D mask after Bayesian polishing showed a strong indication of  $C_2$  symmetry, including the  $M_{\text{lid}}$  modules. Therefore,  $C_2$  symmetry was imposed on the  $C_1$  refined map, which was then compared with the original map. The visual inspection showed that the  $M_{\text{lid}}$  modules were more clearly resolved with the  $C_2$  map. Accordingly,  $C_2$  symmetry was assumed for the subsequent processing to further improve the resolution. Two rounds of 3D classification were performed using the  $C_2$  map, and 358,932 particle images (133 classes) were reselected from

the last round of 2D classification to increase the structural homogeneity of the dataset. A total of 198,918 particles associated with the best class with the highest resolution of the second 3D classification were recentered, re-extracted with a pixel size of 0.83 Å pix<sup>-1</sup>, and then subjected to 3D refinement. After two cycles of CTF refinement<sup>35</sup> and Bayesian polishing<sup>36</sup>, the resulting particles were subjected to two rounds of 3D refinement imposing  $C_2$  symmetry (first no 3D mask and second with a soft-edged 3D mask). Additionally, to investigate the possible asymmetrical nature of substrate binding as well as the  $G_{\omega}$  and  $M_{\text{lid}}$  modules of our research interests, the last two rounds of 3D refinement were repeated assuming asymmetry. Postprocessing yielded a map with a global resolution of 2.52 Å for the  $C_2$  process and 2.64 Å for the  $C_1$  process according to the FSC = 0.143 criteria<sup>37</sup>. The details of the cryo-EM data processing are shown in Supplementary Fig. 4, and the associated parameters and statistics are summarized in Supplementary Table 1. Directional FSC plots were calculated using the 3DFSC server to evaluate the evenness of the distribution of particle orientations (Supplementary Fig. 5b, c)<sup>38</sup>.

### Model building and refinement

Each of the four models of  $G_{\text{core}}$  and  $M_{\text{core}}$  in the *TeCphA1* tetramer was manually built based on the post-processed  $D_2$  cryo-EM map of the apo state using Coot<sup>40</sup>. Model refinement was performed using phenix.real\_space\_refine with the secondary-structure restraints generated by Phenix.secondary\_structure\_restraints in the PHENIX program suite<sup>41,42</sup>. The built models were fit to the  $C_2$  cryo-EM map of the substrate-bound state with UCSF Chimera<sup>43</sup>. The fitted models were readjusted into the map and automatically refined using Coot and PHENIX. The structural models of the other modules, the N domain,  $G_{\text{lid}}$ ,  $G_{\omega}$ , and  $M_{\text{lid}}$ , were generated using the protein structure modeling tools SWISS-MODEL<sup>44</sup> and AlphaFold2<sup>45</sup> to utilize them as initial models for the further model building to generate relatively weak density maps. SWISS-MODEL was run using a part of the sequence of *TeCphA1* (residues 218–481) to build the initial models of  $G_{\text{lid}}$  and  $G_{\omega}$ . The model with the highest GMQE (0.75) and QMEAN (-0.84) scores were obtained from the crystal structure of  $\gamma$ -glutamate-cysteine ligase/glutathione synthetase (GshF, PDB 3LN7), sequence identity of 35.5%) as a template. AlphaFold2 predicted the initial models of the N domain (residues 1–161) and  $M_{\text{lid}}$  (724–879), in which almost the entire region showed high predicted local distance test (pLDDT) scores (>90), except for the terminal residues (residues 160–161, 724–733, and 877–879) and loop regions (residues 58–65 and 826–833). The initial models were fit to the  $C_2$  cryo-EM map of the substrate-bound state with UCSF Chimera. Manual model rebuilding and refinement were performed using Coot and PHENIX, respectively. The refined models were applied to the  $C_1$  cryo-EM maps of the substrate-bound state and  $D_2$  maps of the apo and ATP $\gamma$ S-bound states. After fitting each module to the maps with UCSF Chimera, the models were readjusted to the map and refined using Coot and PHENIX. Finally, ligands were manually fit in the map of the ATP $\gamma$ S-bound and substrate-bound states. The PDB coordinates AGS and 7ID were used for the models of ATP $\gamma$ S and a  $\beta$ -Asp-Arg dipeptide. All the final models were separately validated using MolProbity<sup>46</sup> in PHENIX. The statistics of the model refinement are summarized in Supplementary Table 1. Figures were generated using UCSF Chimera and PyMol (Schrödinger).

### CphA1 activity assays

The reaction mixture contained 2.5–10  $\mu$ g of purified *TeCphA1*, 50 mM Tris-HCl (pH 8.2), 20 mM KCl, 20 mM MgCl<sub>2</sub>, 1 mM DTT, 4 mM ATP, 5 mM L-aspartate, 0.5 mM L-arginine, and 2 mM synthetic cyanophycin primer peptide of different lengths (1- to 5-mer). These are essentially the same conditions as in a previous report<sup>9</sup> with a small modification (replacement of 10 mM 2-mercaptoethanol with 1 mM DTT). For enzyme kinetics, the activity of *TeCphA1* (1  $\mu$ M as a tetramer) was measured by varying the concentration of one of the four substrates,

ATP, L-aspartate, L-arginine, and ( $\beta$ -Asp-Arg)<sub>4</sub>. Kinetic analysis revealed that our enzyme-assay conditions for arginine, ( $\beta$ -Asp-Arg)<sub>4</sub>, and ATP were appropriate. While the concentration of aspartate is rather low for a stable enzyme-activity measurement, we determined the aspartate concentration at 5 mM by referring to an earlier experiment<sup>49</sup>. In the analysis of *TeCphA1* mutants, a 4-mer peptide, ( $\beta$ -Asp-Arg)<sub>4</sub>, was used as a primer. The reaction was performed at 30°C for 30 min, with a total reaction volume of 25  $\mu$ L. To confirm the *TeCphA1* activity, the amount of released phosphate was measured by the colorimetric molybdenum blue method<sup>47</sup>. The postreaction solution was mixed with 250  $\mu$ L of molybdate solution (0.5% ammonium molybdate and 0.5 M H<sub>2</sub>SO<sub>4</sub>) and 50  $\mu$ L of SnCl<sub>2</sub> solution (2 mg ml<sup>-1</sup> SnCl<sub>2</sub> dissolved in HCl and 0.5 M H<sub>2</sub>SO<sub>4</sub>). After incubation for 6 min at 25°C, the absorbance at 660 nm was measured using a Shimadzu UV-1800 spectrometer (Shimadzu Corporation). Data were analyzed using GraphPad Prism (GraphPad Software). In the steady-state kinetics, the data for ATP, L-arginine, and ( $\beta$ -Asp-Arg)<sub>4</sub> were fitted with a Michaelis-Menten curve, and the curve fitting for L-aspartate was performed using an allosteric sigmoidal equation as follows:

$$\text{Reaction rate} = \frac{V_{\max}[S]^h}{K_{\text{half}}^h + [S]^h} = \frac{V_{\max}[S]^h}{K_{\text{prime}} + [S]^h} \quad (1)$$

where  $V_{\max}$  is the maximum enzyme velocity;  $[S]$  is the substrate concentration;  $K_{\text{half}}$  is the concentration of substrate that produces a half-maximal enzyme velocity, and  $h$  is the Hill coefficient<sup>48</sup>.

### Sequence alignments

A homologous sequence search was performed using the amino acid sequences of *TeCphA1* (Accession No. MBS9770029.1) and GshF from *Pasteurella multocida* (Accession No. WP\_010906990.1) in Protein BLAST<sup>49</sup>. CLUSTAL W<sup>50</sup> was used for multiple sequence alignments using default parameters, and the results were displayed by ESPript 3.0<sup>51</sup>. Sequence logos were created using multiple sequence alignment data in the WebLogo server<sup>52</sup>.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

Cryo-EM maps generated in this study have been deposited in the Electron Microscopy Data Bank (EMDB) under accession codes [EMD-32381](#) for *TeCphA1* in the apo state, [EMD-32382](#) for *TeCphA1* bound with ATP $\gamma$ S, and [EMD-32383](#) ( $C_1$ ) and [EMD-32384](#) ( $C_2$ ) for *TeCphA1* bound with ATP $\gamma$ S, aspartate, and ( $\beta$ -Asp-Arg)<sub>4</sub>. The structural coordinates are available in the PDB under accession codes [7WAC](#) for *TeCphA1* in the apo state, [7WAD](#) for *TeCphA1* bound with ATP $\gamma$ S, and [7WAE](#) ( $C_1$ ) and [7WAF](#) ( $C_2$ ) for *TeCphA1* bound with ATP $\gamma$ S, aspartate, and ( $\beta$ -Asp-Arg)<sub>4</sub>. Enzymatic activity data generated in this study are provided in the Source Data file. The PDB coordinates used in this study are as follows: [7LGJ](#) and [7LGQ](#). Source data are provided with this paper.

### References

- Moradali, M. F. & Rehm, B. H. A. Bacterial biopolymers: from pathogenesis to advanced materials. *Nat. Rev. Microbiol.* **18**, 195–210 (2020).
- Rehm, B. H. A. Bacterial polymers: biosynthesis, modifications and applications. *Nat. Rev. Microbiol.* **8**, 578–592 (2010).
- Cai, S. F. et al. A novel DNA-binding protein, PhaR, plays a central role in the regulation of polyhydroxyalkanoate accumulation and granule formation in the haloarchaeon *Haloferax mediterranei*. *Appl. Environ. Microb.* **81**, 373–385 (2015).
- Frommeyer, M., Wiefel, L. & Steinbüchel, A. Features of the biotechnologically relevant polyamide family “cyanophycins” and their biosynthesis in prokaryotes and eukaryotes. *Crit. Rev. Biotechnol.* **36**, 153–164 (2016).
- Du, J. K., Li, L. & Zhou, S. N. Microbial production of cyanophycin: From enzymes to biopolymers. *Biotechnol. Adv.* **37**, 107400 (2019).
- Numata, K. Poly(amino acid)s/polypeptides as potential functional and structural materials. *Polym. J.* **47**, 537–545 (2015).
- Simon, R. D. & Weathers, P. Determination of the structure of the novel polypeptide containing aspartic acid and arginine which is found in cyanobacteria. *Biochim. Biophys. Acta, Protein Struct.* **420**, 165–176 (1976).
- Hoffman, B. M. et al. Mechanism of nitrogen fixation by nitrogenase: the next stage. *Chem. Rev.* **114**, 4041–4062 (2014).
- Burnat, M., Herrero, A. & Flores, E. Compartmentalized cyanophycin metabolism in the diazotrophic filaments of a heterocyst-forming cyanobacterium. *Proc. Natl Acad. Sci. USA.* **111**, 3823–3828 (2014).
- Finzi-Hart, J. A. et al. Fixation and fate of C and N in the cyanobacterium *Trichodesmium* using nanometer-scale secondary ion mass spectrometry. *Proc. Natl Acad. Sci. USA.* **106**, 6345–6350 (2009).
- Watzer, B. & Forchhammer, K. Cyanophycin synthesis optimizes nitrogen utilization in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Appl. Environ. Microb.* **84**, e01298–18 (2018).
- Aboulmagd, E., Sanio, F. B. O. & Steinbüchel, A. Purification of *Synechocystis* sp. strain PCC6308 cyanophycin synthetase and its characterization with respect to substrate and primer specificity. *Appl. Environ. Microb.* **67**, 2176–2182 (2001).
- Sharon, I. et al. Structures and function of the amino acid polymerase cyanophycin synthetase. *Nat. Chem. Biol.* **17**, 1101–1110 (2021).
- Capone, D. G. et al. *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
- Smith, C. A. Structure, function and dynamics in the mur family of bacterial cell wall ligases. *J. Mol. Biol.* **362**, 640–655 (2006).
- Deva, T., Baker, E. N., Squire, C. J. & Smith, C. A. Structure of *Escherichia coli* UDP-N-acetylmuramoyl:L-alanine ligase (MurC). *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 1466–1474 (2006).
- Favini-Stabile, S., Contreras-Martel, C., Thielens, N. & Dessen, A. MreB and MurG as scaffolds for the cytoplasmic steps of peptidoglycan biosynthesis. *Environ. Microbiol.* **15**, 3218–3228 (2013).
- Berg, H. et al. Biosynthesis of the cyanobacterial reserve polymer multi-L-arginyl-poly-L-aspartic acid (cyanophycin): mechanism of the cyanophycin synthetase reaction studied with synthetic primers. *Eur. J. Biochem.* **267**, 5561–5570 (2000).
- Krehenbrink, M. & Steinbüchel, A. Partial purification and characterization of a non-cyanobacterial cyanophycin synthetase from *Acinetobacter calcoaceticus* strain ADP1 with regard to substrate specificity, substrate affinity and binding to cyanophycin. *Microbiol. (Read.)* **150**, 2599–2608 (2004).
- Fawaz, M. V., Topper, M. E. & Firestine, S. M. The ATP-grasp enzymes. *Bioorg. Chem.* **39**, 185–191 (2011).
- Hara, T., Kato, H., Katsube, Y. & Oda, J. A pseudo-michaelis quaternary complex in the reverse reaction of a ligase: structure of *Escherichia coli* B glutathione synthetase complexed with ADP, glutathione, and sulfate at 2.0 Å resolution. *Biochemistry* **35**, 11967–11974 (1996).
- Li, H., Fast, W. & Benkovic, S. J. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. *Protein Sci.* **18**, 881–892 (2009).
- Mol, C. D. et al. Crystal structures of active fully assembled substrate- and product-bound complexes of UDP-N-acetylmuramic acid:L-alanine ligase (MurC) from *Haemophilus influenzae*. *J. Bacteriol.* **185**, 4152–4162 (2003).

24. Šink, R. et al. MurD enzymes: some recent developments. *Biomol. Concepts* **4**, 539–556 (2013).
25. Stout, J. S., De Vos, D., Vergauwen, B. & Savvides, S. N. Glutathione biosynthesis in bacteria by bifunctional gshf is driven by a modular structure featuring a novel hybrid ATP-grasp fold. *J. Mol. Biol.* **416**, 486–494 (2012).
26. Arai, T. & Kino, K. A cyanophycin synthetase from *Thermosynechococcus elongatus* BP-1 catalyzes primer-independent cyanophycin synthesis. *Appl. Microbiol. Biotechnol.* **81**, 69–78 (2008).
27. Zhang, H. et al. The cyanobacterial ornithine–ammonia cycle involves an arginine dihydrolase. *Nat. Chem. Biol.* **14**, 575–581 (2018).
28. Ziegler, K. et al. Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *Eur. J. Biochem.* **254**, 154–159 (1998).
29. Klemke, F. et al. CphA2 is a novel type of cyanophycin synthetase in N<sub>2</sub>-fixing cyanobacteria. *Microbiology* **162**, 526–536 (2016).
30. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
31. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
32. Zhang, K. Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
33. Moriya, T. et al. High-resolution Single Particle analysis from Electron Cryo-microscopy Images Using SPHIRE. *J. Vis. Exp.* **123**, e55448 (2017).
34. Wagner, T. et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. *Commun. Biol.* **2**, 218 (2019).
35. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
36. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
37. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003).
38. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
39. Rohou, A. & Grigorieff, N. CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Bio.* **192**, 216–221 (2015).
40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–221 (2010).
42. Wang, L. et al. Real-space quantum-based refinement for cryo-EM: QR#3. *Acta Crystallogr. D. Biol. Crystallogr.* **76**, 1184–1191 (2020).
43. Pettersen, E. P. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296–W303 (2018).
45. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).
46. Chen, V. B. et al. Molproviy: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).
47. Nagui, E. A., McKelvie, I. D., Worsfold, P. & Kolev, S. D. The molybdenum blue reaction for the determination of orthophosphate revisited: opening the black box. *Anal. Chim. Acta* **890**, 60–82 (2015).
48. Copeland, R. A. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis, 2nd Edition* (Wiley, Hoboken, 2000).
49. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
50. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
51. Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
52. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

## Acknowledgements

We thank Y. Sakamaki and M. Kikkawa (Cryo-EM facility at the University of Tokyo) for their help in cryo-EM data collection and A. Nakamura (Gakusyuin University) for the technical support to build the structural model of (β-Asp-Arg)<sub>4</sub>. This research was supported by the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP21am0101077/JP22ama121010 (T. Miyakawa) and JP21am0101071/JP22ama121001 (support number 1678) (T.S.), and the Finance Science and Technology Project of Hainan Province under Grant Number ZDKJ202018.

## Author contributions

M.T. and T.S. supervised the project. T.Miyakawa, J.Y., and M.T. designed the experiments. J.Y., A.F., and Y.M. prepared recombinant proteins. A.F. and Y.M. performed biochemical studies with assistance from T.Miyakawa. T.Miyakawa, N.A., and M.K. collected cryo-EM data. T.Miyakawa and M.K. processed the data with the support of N.A., T.Moriya, and T.Muramatsu. T.Miyakawa and A.F. performed model building and refinement. T.Miyakawa, J.Y., M.K., N.A., A.F., and T.Moriya wrote the manuscript with assistance from all the authors. M.T. and T.S. edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-32834-8>.

**Correspondence** and requests for materials should be addressed to Toshiya Senda or Masaru Tanokura.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022